Predicting the Focus of Negation: Model and Error Analysis

Md Mosharaf Hossain[*] **Kathleen Hamilton**[**] **Alexis Palmer**[**] **Eduardo Blanco**[*]

[*]Department of Computer Science and Engineering, University of North Texas
mdmosharafhossain@my.unt.edu, eduardo.blanco@unt.edu
[**]Department of Linguistics, University of North Texas
kathleenhamilton@my.unt.edu, alexis.palmer@unt.edu

Abstract

The focus of a negation is the set of tokens intended to be negated, and a key component for revealing affirmative alternatives to negated utterances. In this paper, we experiment with neural networks to predict the focus of negation. Our main novelty is leveraging a scope detector to introduce the scope of negation as an additional input to the network. Experimental results show that doing so obtains the best results to date. Additionally, we perform a detailed error analysis providing insights into the main error categories, and analyze errors depending on whether the model takes into account scope and context information.

1 Introduction

Negation is a complex phenomenon present in all human languages. Horn (2010) put it beautifully when he wrote "negation is what makes us human, imbuing us with the capacity to deny, to contradict, to misrepresent, to lie, and to convey irony." Broadly speaking, negation "relates an expression *e* to another expression with a meaning that is in some way opposed to the meaning of *e*" (Horn and Wansing, 2017). The key challenge to understanding negation is thus to figure out the meaning that is *in some way* opposed to *e*—a semantic and highly ambiguous undertaking that comes naturally to humans in everyday communication.

Negation is generally understood to carry positive meaning, or in other words, to suggest an affirmative alternative. For example, *John didn't leave the house* implies that *John stayed inside the house*. Hasson and Glucksberg (2006) show that comprehending negation involves considering the representation of affirmative alternatives. While not fully understood, there is evidence that negation involves reduced access to the affirmative mental representation (Djokic et al., 2019). Orenes et al. (2014) provide evidence that humans switch

to the affirmative alternative in binary scenarios (e.g., from *not red* to *green* when processing *The figure could be red or green. The figure is not red*). In such multary scenarios, however, humans keep the negated representation unless the affirmative interpretation is obvious from context (e.g., humans keep *not red* when processing *The figure is red, green, yellow or blue. The figure is not red.*).

From a linguistic perspective, negation is understood in terms of scope and focus (Section 2). The scope is the part of the meaning that is negated, and the focus is the part of the scope that is most prominently or explicitly negated (Huddleston and Pullum, 2002). Identifying the focus is a semantic task, and it is critical for revealing implicit affirmative alternatives. Indeed, the focus of negation usually contains only a few tokens, and it is rarely grammatically modified by a negation cue such as *never* or *not*. Only the focus of a negation is actually intended to be negated, and the resulting affirmative alternatives range from implicatures to entailments as exemplified below (focus is underlined, and affirmative alternatives are in italics):

- He didn't report the incident to his superiors <u>until confronted with the evidence</u>.
  He reported the incident to his superiors, but not until confronted with the evidence.
- The board didn't learn <u>the details</u> about the millions of dollars wasted in duplicate work.
  The board learnt about the millions of dollars wasted in duplicate work, but not the details.

In this paper, we experiment with neural networks for predicting the focus of negation. We work with the largest corpus annotating the focus of negation (PB-FOC, 3,544 negations), and obtain the best results to date. The main contributions of this paper are: (a) neural network architecture taking into account the scope of negation and context, (b) experimental results showing that scope information as predicted by an automated scope detector is more

beneficial than context, (c) quantitative analysis profiling which foci are easier and harder to predict, and (d) detailed qualitative analysis providing insights into the errors made by the models. Crucially, the scope detector we leverage to predict focus is trained with CD-SCO, a corpus created independently of PB-FOC (Section 2). Our results suggest that negation scopes may transfer across (a) genres (short stories vs. news) and (b) negation types (all negations vs. only verbal negations, i.e., when the negation cue modifies a verb).

2 Background

It is generally understood that negation has scope and focus. Scope is "the part of the meaning that is negated" and includes all elements whose individual falsity would make the negated statement strictly true (Huddleston and Pullum, 2002). Consider the following statement (1) *John doesn't know exactly how they met*. This statement is true if one or more of the following propositions are false: (1a) Somebody *knows* something, (1b) *John* is the one who knows, (1c) *exactly* is the manner of *knowing*, and (1d) *how they met* is what is known. Thus, the scope of the negation in statement (1) is (1a–d). The focus of a negation is "the part of the scope that is most prominently or explicitly negated", or in other words, the element of the scope that is intended to be interpreted as false to make the overall negative true (Huddleston and Pullum, 2002). Determining the focus consists in pinpointing which parts of the scope are intended to be interpreted as true and false given the original statement. Without further context, one can conclude that the intended meaning of statement (1) is *John knows how they met, but not exactly*, or alternatively, that (1a–b, 1d) are intended to be interpreted as true, and (1c) as false. This interpretation results from selecting as focus (1c), i.e., the manner of *knowing*.

We summarize below corpora annotating scope and focus of negation, emphasizing the ones we work with. The survey by Jiménez-Zafra et al. (2020) provides a more comprehensive analysis including corpora in languages other than English.
Corpus Annotating Scope. In the experiments described here, we work with a scope detector trained with CD-SCO (Morante and Daelemans, 2012), which annotates negation cues and negation scopes in two stories by Conan Doyle: The Hound of the Baskervilles and The Adventure of Wisteria Lodge. The corpus contains 5,520 sentences, 1,227

	%foci	%verb with	%role is focus
ARG$_0$	4.09	67.44	6.06
ARG$_1$	43.76	90.47	48.36
ARG$_2$	5.53	14.24	38.81
ARG$_3$	0.39	1.49	26.42
ARG$_4$	0.51	0.79	64.29
M-NEG	26.08	99.89	26.11
M-TMP	7.16	16.80	42.62
M-MNR	5.50	7.36	74.71
M-ADV	3.30	13.53	24.38
M-LOC	1.01	3.72	27.27
M-EXT	0.45	0.56	80.00
M-DIR	0.25	1.07	23.68
M-PNC	1.49	2.42	61.63
M-DIS	0.28	7.81	3.61
M-CAU	0.11	2.88	3.92

Table 1: Analysis of PB-FOC: overall percentages of foci per role, percentages of negated verbs having each role, and percentage of each role being the focus.

of which contain a negation. CD-SCO annotates all negations, including verbs (e.g., I *fail* to see how you could have done more), adverbs (e.g., It was *never* proved that [. . .]), determiners (e.g., There is *no* friend like [. . .]), pronouns (e.g., [. . .] has yielded *nothing* to a careful search), affixes (e.g., The *in*explicable tangle seemed [. . .]), and others.

Other corpora annotating scope in English include efforts with biomedical texts (Vincze et al., 2008) and working with reviews (Councill et al., 2010; Konstantinova et al., 2012).
Corpora Annotating Focus. Although focus of negation is defined as a subset of the scope, there is no corpus annotating both of them in the same texts. We work with PB-FOC, the largest publicly available corpus annotating focus of negation (Blanco and Moldovan, 2011). PB-FOC annotates the focus of the negations marked with M-NEG role in Prop-Bank (Palmer et al., 2005), which in turn annotates semantic roles on top of the Penn TreeBank (Taylor et al., 2003). As a result, PB-FOC annotates the focus of 3,544 verbal negations (i.e., when a negation cue such as *never* or *not* syntactically modifies a verb). As per the authors, the annotation process consisted of selecting the semantic role most likely to be the focus. Therefore, focus annotations in PB-FOC are always all the tokens corresponding to a semantic role of the (negated) verb. Finally, M-NEG role is chosen when the focus is the verb. The annotations in PB-FOC were carried out taking into account the previous and next sentences. We provide examples below, and Section 5 provides ad-

ditional examples. We indicate the semantic roles in PropBank with square brackets, and the role selected as focus is underlined.

- Even if [that deal]$_{\text{ARG}_1}$ is[n't]$_{\text{M-NEG}}$ [revived]$_{verb}$, NBC hopes to find another.
- [A decision]$_{\text{ARG}_1}$ is[n't]$_{\text{M-NEG}}$ [expected]$_{verb}$ [until some time next year]$_{\text{M-TMP}}$.
- But [quite a few money managers]$_{\text{ARG}_0}$ are[n't]$_{\text{M-NEG}}$ [buying]$_{verb}$ [it]$_{\text{ARG}_1}$.

Table 1 presents basic statistics for PB-FOC. ARG_1 is the most frequent role to be focus (43.76%) followed by M-NEG (26.08%) and a relatively long list of infrequent roles (ARG_0, ARG_2, M-TMP, M-MNR: 4.09–7.16%). More interestingly, the last two columns in Table 1 indicate (a) how often a negated verb has each semantic role, and (b) how often a role of a negated verb is the focus—if a negated verb-argument structure does not have a particular role, that role obviously cannot be the focus. These percentages reveal that role presence does not uniquely identify foci, but some semantic roles, although infrequent overall, are likely to be the focus if present (M-EXT: 80.00%, M-MNR: 74.71%, ARG_4: 64.29%, M-PNC: 61.63%).

Other corpora annotating the focus in English redefine the annotation guidelines (Anand and Martell, 2012), use dependency trees instead of roles (Sarabi and Blanco, 2016), target non-verbal negations (Sarabi and Blanco, 2017), and work with tutorial dialogues (Banjade and Rus, 2016).

3 Previous Work

In addition to identifying negation cues and resolving the scope and focus of negation, there is work showing that processing negation is important for natural language understanding in general. In particular, sentiment analysis benefits from processing negation (Wiegand et al., 2010). For example, *like* generally carries positive sentiment, but not when modified by a negation cue (e.g., *don't like*). Wilson et al. (2005) introduce the idea of contextual polarity, and note that negation may intensify rather than change polarity (e.g., *not good* vs. *not only good but amazing*). Jia et al. (2009) present a set of heuristic rules to determine sentiment when negation is present, and Councill et al. (2010) show that information about the scope of negation is beneficial to predict sentiment. Outside sentiment analysis, Bentivogli et al. (2016) point out that neural machine translation struggles translating negation, and point to focus detection as a possible solution.

Neural networks are hard to interpret, but there is evidence that they learn to process negation—to a certain degree—when trained to predict sentiment analysis. Li et al. (2016) visually show that neural networks are capable of meaning composition in the presence of, among others, negation and intensification. Wang et al. (2015) show that an LSTM architecture is capable of determining sentiment of sequences containing negation such as *not good* and *not bad*. These previous works train a model for a particular task (i.e., sentiment analysis) and then investigate whether the model learnt anything related to negation that is useful for that task. Unlike them, we target focus of negation detection—and the resulting affirmative alternatives—and work with task-independent negations.

Scope Identification. Compared to focus identification, scope identification has received substantially more attention. The first proposals (Morante and Daelemans, 2009) were trained in the biomedical domain with BioScope (Szarvas et al., 2008). The *SEM-2012 Shared Task (Morante and Blanco, 2012) included scope identification with CD-SCO (Section 2), and the winner proposed an SVM-based ranking of syntactic constituents to identify the scope (Read et al., 2012). More recently, Fancellu et al. (2016) present neural networks for this task, and Packard et al. (2014) present a complementary approach that operates over semantic representations obtained with an off-the-shelf parser. Finally, Fancellu et al. (2017) present an error analysis showing that scope is much easier to identify when delimited by punctuation. In this paper, we use a scope detector trained with CD-SCO to predict the focus of negation. While we only incorporate small modifications to previously proposed architectures, our scope detector outperforms previous work (Section 4).

Focus Identification. Although focus is part of the scope, state-of-the-art approaches to identify the focus of negation ignore information about scope. Possible reasons are that (a) existing corpora annotating scope and focus contain substantially different texts (Section 2), and (b) incorporating scope information is not straightforward with traditional machine learning and manually defined features. The initial proposals obtain modest results and only consider the sentence containing the negation (Blanco and Moldovan, 2011), including scope information in a rule-based system (Rosenberg and Bergler, 2012). Zou et al. (2014, 2015) propose
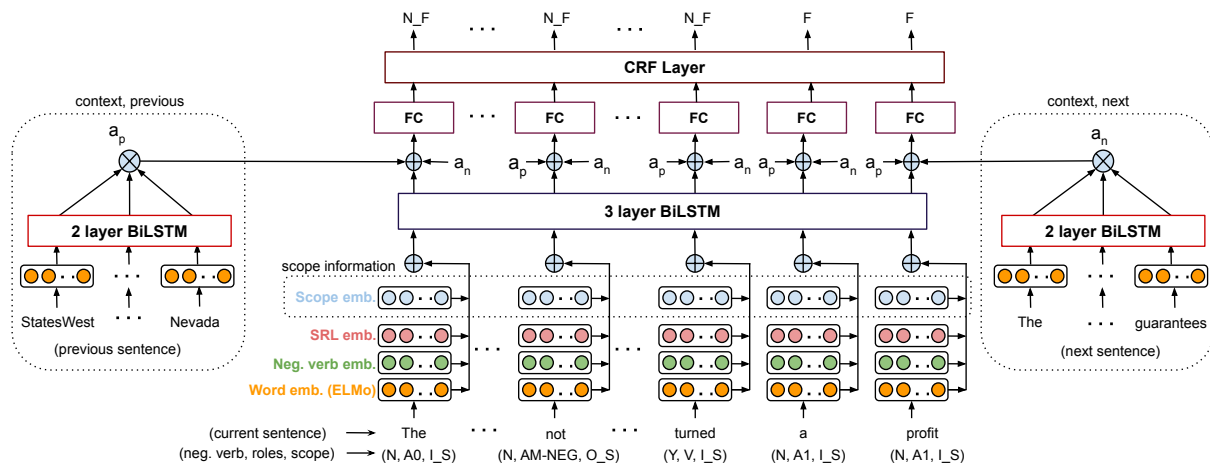
Figure 1: Neural network to predict the focus of negation. The core of the architecture (NN, all components except those inside dotted shapes) takes as input the sentence containing the negation, and each word is represented with its word embedding and specialized embeddings for the negated verb and semantic roles. The additional components inside dotted shapes incorporate information about (a) the scope and (b) context (previous and next sentences).

graph-based models that incorporate discourse information and obtain improvements over previous works. In addition, Shen et al. (2019) present a neural model that leverages word-level and topic-level attention mechanisms to utilize contextual information. We compare our results and theirs in Section 4.2. In this paper, we show that (a) neural networks considering the scope of negation obtain the best results to date and (b) context is not beneficial if scope is available (Section 4).

4 Predicting the Focus of Negation

We approach the task of predicting focus of negation as a sequence labeling task with a neural network. We first describe the network architecture, and then present quantitative results. Section 5 presents a detailed error and qualitative analysis.

4.1 Neural Network Architecture

The network architecture (Fig. 1) consists of a base NN (all components except those inside dotted shapes) plus additional components to include information about the scope and context of negation.
Base NN. The base network is inspired by Huang et al. (2015) and Reimers and Gurevych (2017). It is a 3-layer Bidirectional Long Short-Term Memory (BiLSTM) network with a Conditional Random Field (CRF) layer. The network takes as input the sentence containing the negation whose focus is to be predicted, where each word is represented with the concatenation of (a) its pre-trained ELMo embedding Peters et al. (2018), (b) a specialized em-

bedding indicating whether a token is the negated verb (not the negation cue), and (c) a specialized embedding indicating semantic roles (one per role label). The specialized embeddings are trained from scratch as part of the tuning of the network.
Scope Information. We add an extra input at the token level indicating whether a token belongs to the scope of the negation whose focus is to be predicted. This new input is then mapped to a third specialized embedding (two values: inside or outside the scope), and concatenated to the word representation prior to feeding it to the 3-layer BiLSTM.

Scope information is taken from a scope detector inspired by Fancellu et al. (2016). Our modifications are as follows. First, we add a CRF layer on top of the 2-layer BiLSTM. Second, we use GloVe embeddings instead of word2vec embeddings. We train the scope detector with CD-SCO (Section 3), and our simple modifications yield the best results to date predicting the scope of negation: 79.41 F1 (vs. 77.77 F1). We do not elaborate on the scope detector as we only leverage it to predict focus.
Context. We also experiment with an additional component to add contextual information (previous and next sentences), as previous work has shown empirically that doing so is beneficial (Zou et al., 2014). While we tried many strategies (e.g., concatenating sentence embeddings to the representations from the 3-layer BiLSTM), we present only the one yielding the best results. Specifically, we use 2-layer Bi-LSTMs with an attention mechanism (Bahdanau et al., 2014; Yang et al., 2016). The attention weights (a_p and a_n for the previous

	P	R	F1	Acc
Zou et al. (2014)	71.67	67.43	69.49	67.1
Zou et al. (2015)	n/a	n/a	n/a	69.4
Shen et al. (2019)	n/a	n/a	n/a	70.5
NN (baseline)	72.14	71.63	71.88	71.6
NN + S	**75.92**	**75.7**	**75.81**	**75.7**
NN + Cntxt	73.69	73.17	73.43	73.2
NN + S + Cntxt	74.15	73.74	73.94	73.7

Table 2: Focus prediction results of the best performing previous works and our neural network (baseline network and adding components). S and Cntxt refer to Scope and Context, respectively. Note that Zou et al. (2014) do not report the accuracy of their model, but they do in their follow-up work (Zou et al., 2015).

	%insts.	P	R	F1
ARG$_0$	4.07	92.9	44.8	60.5
ARG$_1$	43.82	77.9	90.4	83.7
ARG$_2$	4.92	62.0	88.6	72.9
ARG$_3$	0.42	16.7	33.3	22.2
ARG$_4$	0.56	60.0	75.0	66.7
M-NEG	25.98	83.8	50.3	62.8
M-TMP	7.16	71.2	92.2	80.3
M-MNR	5.76	88.6	95.1	91.8
M-ADV	3.09	85.0	77.3	81.0
M-LOC	1.12	60.0	75.0	66.7
M-EXT	0.84	100.0	100.0	100.0
M-DIR	0.28	50.0	100.0	66.7
M-PNC	1.69	84.6	91.7	88.0
M-DIS	0.14	100.0	100.0	100.0
M-CAU	0.14	0.0	0.0	0.0

Table 3: Results per role with our best system (NN + Scope, Figure 1). % insts. indicates the percentage of foci per role in the test set.

and next sentences respectively) are concatenated to the representations from the 3-layer BiLSTM.

Hyperparameters and Training Details. The cell states of all BiLSTMs have size 350 and we use dropout with a ratio of 0.6. We use the stochastic gradient descent algorithm with Adam optimizer (Kingma and Ba, 2014) and a learning rate of 0.001 for tuning weights. We set batch size to 24 and stop the training process after the F1 on the development split does not increase for 50 epochs. The final model is the one which yields the highest F1 on the development split. We combined the original train and development splits from PB-FOC and used 95% of the result as training split and the remaining 5% as development split. The implementation uses PyTorch (Paszke et al., 2019).[1]

We refer the readers to the supplemental material for additional details on the neural architecture.

4.2 Quantitative Analysis

Table 2 presents the results obtained with the *SEM Shared Task test split and evaluation script. Our best network architecture (NN + Scope) outperforms all previous works (Accuracy: +5.2, 7.4%).

Not all components of the architecture we experiment with are beneficial. Our main finding is that scope information, as predicted by a scope detector trained on CD-SCO, is very useful. Indeed, the core of the network (3-layer BiLSTM and CRF layer) obtains 75.81 F1 (vs. 71.88) when the input includes scope information. Disabling other specialized embeddings—indicating the negated verb and semantic roles—results in substantial drops in performance (not shown in Table 2).

According to the creators of PB-FOC and more recent work (Zou et al., 2014, 2015), context is important to determine the focus of negation. Our results confirm this observation: adding the previous and next sentences via attention mechanisms improves the results: 73.43 vs. 71.88 F1. Our results also show, however, that the scope of negation—not previously considered—is more beneficial than context. As a matter of fact, adding context is detrimental if scope is taken into account.

Table 3 presents the results of the best system (NN + Scope) per role. We observe that all roles obtain relatively high F1 scores (>60.5) with two exceptions: ARG$_3$ (22.2) and M-CAU (0.0). Many roles are rarely the focus ($\leq$5%: ARG$_0$, ARG$_2$, ARG$_3$, ARG$_4$, etc.), yet the F1 scores with those roles are similar or even higher than more frequent roles (e.g., ARG$_1$). In other words, the neural model is able to predict the focus with similar F1 scores, regardless of what role is the focus.

In Table 4, we provide a quantitative analysis of the results obtained with the best system (NN + Scope). We split the test set into four categories and subcategories, and then evaluate the test instances that fall into each subcategory. Specifically, we consider the focus length measured in tokens, the sentence length measured in tokens, the number of roles in the verb-argument structure of the negated verb (intuitively, the more roles to choose from, the harder to predict the right one), and the verb class of the negated verb. We obtained verb classes from the lexical files in WordNet (Miller, 1995).

---

[1] Code available at https://github.com/mosharafhossain/focus-of-negation

		%insts.	P	R	F1
focus length	1	39.47	85.2	61.2	66.0
	2–5	33.85	92.2	85.5	87.7
	6–15	21.91	95.3	93.6	93.7
	>15	4.78	89.7	82.4	84.1
sent. length	5–10	7.44	82.6	73.6	74.1
	11–15	10.39	88.0	85.1	85.5
	16–30	46.63	79.8	77.7	76.7
	>30	35.53	77.9	75.9	75.1
#roles	2, 3 roles	10.81	90.3	89.6	89.7
	4 roles	35.25	80.4	79.3	77.2
	5 roles	37.50	77.8	77.5	76.1
	>5 roles	16.43	72.9	65.8	64.7
verb class	possession	17.70	75.1	73.0	71.0
	commun.	14.04	80.0	80.0	79.7
	cognition	12.36	88.9	85.2	81.8
	social	10.81	77.2	75.3	74.3

Table 4: Quantitative analysis of the results in the test set. We measure focus and sentence lengths in tokens. We provide weighted averages per label, thus the F1 scores may not fall between P and R.

Regarding focus length, we observe that single-word foci are the hardest followed by long foci (over 15 tokens). This leads to the conclusion that the network struggles to represent single words and long sequences of words. We note that many foci are single words (39.47%) despite this subcategory obtaining the worst results (F1: 66.0). Regarding sentence length, we observe comparable F1 scores (74.1–76.7) except with sentences between 11 and 15 tokens (85.5). These results lead to the conclusion that since the focus prediction task is defined at the semantic role level, role length is more important than sentence length.

Unsurprisingly, the model obtains worse results depending on the number of roles in the verb-argument structure of the negated verb—effectively, the model suffers when it has more roles to choose from. Negated verbs with up to three roles obtain the highest F1 scores (89.7), and results drop significantly (64.7) when there are more than 5 roles (only 16.43% of instances).

Finally, we provide detailed results for the verbs belonging to the most frequent verb classes: possession (*buy*, *take*, *get*, etc.), communication (*say*, *allege*, etc.), cognition (*think*, *believe*, *imagine*, etc.), and social (*meet*, *party*, etc.). Communication and cognition verbs obtain the best results; this is due in part to the fact that verbs belonging to those verb classes tend to have fewer semantic roles.

5 Error and Qualitative Analysis

To better understand the strengths and weaknesses of our models, we perform a detailed qualitative analysis of the errors made in predicting focus. Negation is a complex semantic phenomenon which interacts with other aspects of the meaning and structure of sentences, and this complexity is reflected in the diversity of errors. We perform the analysis over all 712 negations in the test set, investigating how linguistic properties of the negated sentences influence performance across the four models (baseline, scope, context, and combined); we consider nearly 3,000 predictions in total. The counts in this section reflect instance-model pairings; it could happen, for example, that three of the four models predict the wrong focus for a sentence with a particular linguistic property. For some sentences, multiple error types are relevant.

We identify three broad categories of errors: syntactic (5.1), semantic (5.2), and other (5.3). There are multiple error types within each category, and each error type is associated with a particular linguistic property of the negated sentence. Here we focus on the most frequently-occurring error types per category, as these offer the greatest insight into specific strengths and weaknesses of the models.

The distribution of error categories across the four models is shown in Table 8 and discussed in more detail below (5.4).

Representative examples from PB-FOC for each error type appear in Tables 5, 6, and 7. For each example, we show the full sentence, with predicted scope (as output by the scope detector trained with CD-SCO) between double angle brackets and semantic roles in square brackets. For each negated sentence, the table shows the gold focus (**GF**)[2] and the predicted focus (**PF**), along with the model(s) responsible for the incorrect prediction.

5.1 Syntactic Error Types

Our analysis reveals three prominent error types related to the structure of negated sentences.

1. Complex verb errors occur when the target verb is part of a complex verb constellation, due to passivization, complex tense constructions, or modal constructions. These constructions result in multi-word verb constellations, such as *can't be cured* in example 1.1 (Table 5). These are challeng-

---

[2]Gold focus annotations come from the PB-FOC corpus and may include some errors. Some properties of the PB-FOC annotations are discussed in Section 2.

Syntactic Error Type	Examples from PB-FOC
1.1. Complex verb (all models) ————→	There is [nothing]$_{\text{ARG}_2}$ wrong with the market ≪ [that]$_{\text{ARG}_2}$ [ca]$_{\text{M-MOD}}$ n't be [cured]$_{verb}$ [by a little coherence and common sense in Washington.]$_{\text{ARG}_3}$ ≫
	GF: [*nothing ... that*]$_{\text{ARG}_2}$ **PF:** [*a little coherence and common sense in Washington*]$_{\text{ARG}_3}$
1.2. Complex sentence (NN + context) ————→	Since production costs were guaranteed, it didn't matter that ≪ [a program]$_{\text{ARG}_1}$ [could]$_{\text{M-MOD}}$ n't be [sold]$_{verb}$ [abroad]$_{\text{M-LOC}}$ or put into syndication, ≫ [as most American programs are.]$_{\text{M-ADV}}$
	GF: [*abroad*]$_{\text{M-LOC}}$ **PF:** [*as most American programs are*]$_{\text{M-ADV}}$
1.3. Role adjacency (NN + scope + context) ——→	It was an overreaction to [an event (the failure of a management and union group to get bank financing for a takeover of UAL) that]$_{\text{ARG}_0}$ ≪ doesn't [mean]$_{verb}$ [that much]$_{\text{ARG}_1}$ [to lots of stocks.]$_{\text{M-MNR}}$ ≫
	GF: [*that much*]$_{\text{ARG}_1}$ **PF:** [*much*]$_{\text{ARG}_1}$ [to lots of stocks]$_{\text{M-MNR}}$

Table 5: Syntactic error types. KEY: [semantic role], ≪predicted scope≫, **GF:** gold focus, **PF:** predicted focus.

ing for all models, but especially for the baseline, with 56 error cases (vs. 36, 43, and 41 for the scope, context, and combined models).

2. Complex sentence structure errors are even more common, with 116/73/87/63 occurrences for the four models. Instances triggering this error type are sentences with relative clauses or complement clauses, as well as sentences with non-canonical linking between argument structure and grammatical function, such as passives and questions. According to Horn (2010), relative and complement clauses can alter the behavior of negation, compared to simple declarative sentences. Example 1.2 in Table 5 shows scope helping with complex sentence structure—both models which incorporate scope predict the correct focus, which occurs within the predicted scope. The other two models choose an argument outside of the predicted scope.

Our third type of syntactic error occurs due to **3. Role adjacency** in the sentence, leading to errors in span prediction. The property associated with this error type is linear adjacency of semantic roles, with no textual material in between. Example 1.3 in Table 5 shows that the model predicts part of the correct role but then extends the span to incorporate a second role.

In summary, models with access to predicted scope make fewer syntactic errors than models without scope.

5.2 Semantic Error Types

Three different types of errors related to meaning occur with high frequency.

1. Errors due to distractors are the most fre-

quent individual error type. The term *distractor* is most familiar from pedagogical discussion of multiple-choice questions, where a distractor is an incorrect option that test-takers are likely to mistake for a correct answer. We use the term here to refer to textual material which leads the neural network away from the gold focus. Specifically, distractors are found in two aspects of the input representation for a given instance: the predicted scope, and the adjacent sentences (previous and next) provided as part of the models which incorporate context. This error type is, by definition, not applicable for the baseline model.

We identify 124 occurrences of distractor errors for the scope model, 87 for the context model, and 130 for the combined model, making this the largest error category. Example 2.1 in Table 6 marks distractors in bold-face type. In this case, all models predict *after the last crash* as the focus.[3] The predicted focus occurs in the predicted scope, and the head noun *crash* appears in the surrounding context. In addition to the direct repetition *the 1987 crash* in the sentence following, we see the synonym *market plunge* in the previous sentence.

2. Lack of referential specificity in the gold focus is a less-frequent and more speculative error type. The idea is that focus is difficult to predict correctly when the focused semantic role is pronominal or otherwise requires additional information for reference resolution. Across the models, we count 22 occurrences. In most of these cases, the gold focus is a pronoun (*it*, ex. 2.2). All models seem to

---

[3]An argument could be made for M-NEG as the negated role; however, we show the gold focus according to PB-FOC.

Semantic Error Type	Examples from PB-FOC
2.1. Distractors	A further slide also would resurrect debate over a host of [other, more sweeping changes]$_{ARG_1}$ proposed – but [not]$_{M-NEG}$ $\ll$ [implemented]$_{verb}$ [after the last crash.]$_{M-TMP}$ $\gg$
(all models) $\longrightarrow$	**GF:** [*other, more sweeping changes*]$_{ARG_1}$ **PF:** [*after the last crash*]$_{M-TMP}$
	PrevSent: A deeper **market plunge** today could give them ... **NextSent:** Most notably, several of the regulatory steps recommended by the Brady Task Force, which analyzed **the 1987 crash**, would be revived ...
2.2. Lack of specificity	The main advantage of a convertible mortgage is that $\ll$ [it]$_{ARG_0}$ is $\gg$ not a sale and [therefore]$_{M-DIS}$ $\ll$ does not [trigger]$_{verb}$ [costly transfer taxes and reappraisal.]$_{ARG_1}$ $\gg$
(all models) $\longrightarrow$	**GF:** [*it*]$_{ARG_0}$ **PF:** [*costly transfer taxes and reappraisal*]$_{ARG_1}$
2.3. Neg. Polarity Items (NPIs)	[And]$_{M-DIS}$ [unlike IBM's water-cooled mainframes]$_{M-ADV}$, $\ll$ [it]$_{ARG_0}$ doesn't [need]$_{verb}$ [any plumbing.]$_{ARG_1}$ $\gg$
(all models) $\longrightarrow$	**GF:** [*it*]$_{ARG_0}$ **PF:** [*any plumbing*]$_{ARG_1}$

Table 6: Semantic error types. KEY: [semantic role], $\ll$predicted scope$\gg$, **GF:** gold focus, **PF:** predicted focus.

Other Error Types	Examples from PB-FOC
3.1. Quotations	[No,]$_{M-DIS}$ to my mind, $\ll$ [the Journal]$_{ARG_0}$ did not ["defend]$_{verb}$ [sleaze,]$_{ARG_1}$ $\gg$ [fraud, waste, embezzlement, influence-peddling and abuse of the public trust."]$_{ARG_1}$
(all models) $\longrightarrow$	**GF:** [*not*]$_{M-NEG}$ **PF:** [*sleaze, fraud, waste, ... public trust*]$_{ARG_1}$
3.2. Particle verbs, PPs and inf. complements	[But]$_{M-DIS}$ $\ll$ don't [pay]$_{verb}$ [30 times earnings]$_{ARG_1}$ [for a company that's expected to grow at 15% a year.]$_{ARG_3}$ $\gg$
(all models) $\longrightarrow$	**GF:** [*for a company that's expected to grow at 15% a year*]$_{ARG_3}$ **PF:** [*30 times earnings*]$_{ARG_1}$

Table 7: Other error types. KEY: [semantic role], $\ll$predicted scope$\gg$, **GF:** gold focus, **PF:** predicted focus.

disprefer predicting bare pronouns as focus.

Occurrence of **3. negative polarity items (NPIs)** also influences the accuracy of the model. Negative polarity items (such as *any* or *yet*, see Horn (2010)) are licensed in the scope of negation but ungrammatical elsewhere. For example, it's ungrammatical to say **I have eaten any fish*. Given the strong association between negation and NPIs, it is not surprising that our models tend to predict as focus any role which contains an NPI (example 2.3). This error type occurs roughly twice as often in models with scope than in models without scope.

5.3 Other Error Types.

Two other error types occur often enough to deserve mention. **1. Quotation errors** generally involve quoted direct speech, which seems to be especially problematic when only part of a clause is quoted speech. In example 3.1, the quoted speech is the verb plus its direct object, and all models select the role of the direct object as predicted focus. The final error type is a sort of catch-all: **2. Particle verbs, prepositional phrases, and infinitival complements.** As with complex sentence structures, these error types reflect complex verbal argument structure.

5.4 Discussion

Table 8 shows the distribution of error types across the four systems. Errors due to particular syntactic structures are the most common, with the subtype of **complex sentences** making up the bulk of these (339).[4] The baseline network deals very poorly with both complex verb constellations and complex sentence structures, and incorporating predicted scope consistently reduces the number of errors

---

[4] An error count is incremented whenever the relevant linguistic property is identified in a sentence for which the relevant system has made an incorrect prediction. Note that one sentence may present more than one linguistic property.

	Synt.	Sem.	Other
n	593	407	164
NN (baseline)	32.2	3.0	25.6
NN + Scope	21.3	35.4	24.4
NN + Context	25.6	24.3	25.6
NN + Scope + Context	20.9	37.3	24.4
Total	100.0	100.0	100.0

Table 8: Number of errors made by all systems per category (n), and percentage made by each system.

	Synt.	Sem.	Other
NN (baseline)	78.6	4.5	16.9
NN + Scope	40.5	46.6	12.9
NN + Context	51.7	33.9	14.4
NN + Scope + Ctx	39.4	47.9	12.7

Table 9: Percentages per error category, for each system.

of this type. This suggests that considering scope helps the system to deal with complex sentences.

For errors related to semantics, the picture is reversed. The systems which consider scope are especially prone to **distractor** errors, the most common error type over all (341). When we have both scope and context, the system has even more potential distractor candidates and makes more errors. The two error types in the Other category are distributed roughly evenly across the models, suggesting that none of the current models is any better than the others at dealing with these error types.

In Table 9 we see a second view on the error distributions, now considering each category as a proportion of the errors made by the system. Again we see that predicted scope shifts the balance of error types from syntactic to semantic. By reinforcing a subsection of the text in the input representation, the search space for complex sentences narrows and the system has a better chance of selecting the correct focus. This same behavior is a disadvantage when the gold focus is not part of the predicted scope, as the scope distracts attention away from other plausible candidate roles. Similarly, including context through adjacent sentences sometimes reinforces the correct focus through introduction of other semantically-related terms, and sometimes clutters the field through the very same mechanism.

6 Conclusions

Negation is generally understood to carry positive meaning, or in other words, to suggest affirmative alternatives. Predicting the focus of negation (i.e.,

pinpointing the usually few tokens that are actually negated) is key to revealing affirmative alternatives.

In this paper, we have presented a neural architecture to predict the focus of negation. We work with PB-FOC, a corpus of verbal negations (i.e., when a negation cue grammatically modifies a verb) in which one semantic role is annotated as focus. Experimental results show that incorporating scope of negation information yields better results, despite the fact that we train the scope detector with data in a different domain (short stories vs. news). These results suggest that scope of negation transfers across domains. Our best model (NN + Scope) obtains the best focus prediction results to date. A quantitative analysis shows that this model is robust across most role labels (Table 3), sentence lengths, and verb classes (Table 4). The model obtains worse results, however, when the role that is the focus is only one token, or the negated verb has more than 5 roles (Table 4).

In addition to state-of-the-art results, we have presented a detailed qualitative analysis. We discover three main error categories (syntactic, semantic, and other) and 8 error types after manual analysis of the predictions made by the four models with all test instances. We draw two main insights from the qualitative analysis. First, including scope information solves many syntactic errors but introduces semantic errors (recall that scope information is beneficial from a quantitative point of view). Second, the lower results after including context, at least with the current architecture, are largely due to additional semantic errors via distractors in the previous and next sentences.

Acknowledgements

References

Pranav Anand and Craig Martell. 2012. Annotating the focus of negation in terms of questions under

discussion. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 65–69, Jeju, Republic of Korea. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Rajendra Banjade and Vasile Rus. 2016. Dt-neg: Tutorial dialogues annotated for negation scope and focus in context. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Luisa Bentivogli, Arianna Bisazza, Mauro Cettolo, and Marcello Federico. 2016. Neural versus phrase-based machine translation quality: a case study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 257–267, Austin, Texas. Association for Computational Linguistics.

Eduardo Blanco and Dan Moldovan. 2011. Semantic representation of negation using focus detection. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 581–589, Portland, Oregon, USA. Association for Computational Linguistics.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 152–164, Ann Arbor, Michigan. Association for Computational Linguistics.

Isaac Councill, Ryan McDonald, and Leonid Velikovich. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59, Uppsala, Sweden. University of Antwerp.

Vesna Djokic, Jean Maillard, Luana Bulat, and Ekaterina Shutova. 2019. Modeling affirmative and negated action processing in the brain with lexical and compositional semantic models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5155–5165, Florence, Italy. Association for Computational Linguistics.

Federico Fancellu, Adam Lopez, and Bonnie Webber. 2016. Neural networks for negation scope detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 495–504, Berlin, Germany. Association for Computational Linguistics.

Federico Fancellu, Adam Lopez, Bonnie Webber, and Hangfeng He. 2017. Detecting negation scope is easy, except when it isn't. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 58–63, Valencia, Spain. Association for Computational Linguistics.

Uri Hasson and Sam Glucksberg. 2006. Does understanding negation entail affirmation?: An examination of negated metaphors. *Journal of Pragmatics*, 38(7):1015–1032.

Laurence R Horn, editor. 2010. *The expression of negation*. Mouton de Gruyter.

Laurence R. Horn and Heinrich Wansing. 2017. Negation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, spring 2017 edition. Metaphysics Research Lab, Stanford University.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.

Lifeng Jia, Clement Yu, and Weiyi Meng. 2009. The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1827–1830. ACM.

Salud María Jiménez-Zafra, Roser Morante, María Teresa Martín-Valdivia, and L. Alfonso Ureña-López. 2020. Corpora annotated with negation: An overview. *Computational Linguistics*, 46(1):1–52.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization.

Natalia Konstantinova, Sheila C.M. de Sousa, Noa P. Cruz, Manuel J. Maña, Maite Taboada, and Ruslan Mitkov. 2012. A review corpus annotated for negation, speculation and their scope. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3190–3195, Istanbul, Turkey. European Language Resources Association (ELRA).

Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. 2016. Visualizing and understanding neural models in NLP. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 681–691, San Diego, California. Association for Computational Linguistics.

George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Roser Morante and Eduardo Blanco. 2012. *SEM 2012 Shared Task: Resolving the Scope and Focus of Negation. In *Proceedings of the First Joint Conference on*

Lexical and Computational Semantics (*SEM 2012), pages 265–274, Montréal, Canada.

Roser Morante and Walter Daelemans. 2009. A meta-learning approach to processing the scope of negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 21–29, Boulder, Colorado. Association for Computational Linguistics.

Roser Morante and Walter Daelemans. 2012. ConanDoyle-neg: Annotation of negation cues and their scope in Conan Doyle stories. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. European Language Resources Association (ELRA).

Isabel Orenes, David Beltrán, and Carlos Santamaría. 2014. How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, 74:36–45.

Woodley Packard, Emily M. Bender, Jonathon Read, Stephan Oepen, and Rebecca Dridan. 2014. Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 69–78, Baltimore, Maryland. Association for Computational Linguistics.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proc. of NAACL*.

Jonathon Read, Erik Velldal, Lilja Øvrelid, and Stephan Oepen. 2012. UiO1: Constituent-based discriminative ranking for negation resolution. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 310–318, Montréal, Canada. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2017. Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. *arXiv preprint arXiv:1707.06799*.

Sabine Rosenberg and Sabine Bergler. 2012. UConcordia: CLaC negation focus detection at *Sem 2012. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 294–300, Montréal, Canada. Association for Computational Linguistics.

Zahra Sarabi and Eduardo Blanco. 2016. Understanding negation in positive terms using syntactic dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1108–1118, Austin, Texas. Association for Computational Linguistics.

Zahra Sarabi and Eduardo Blanco. 2017. If no media were allowed inside the venue, was anybody allowed? In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 860–869, Valencia, Spain. Association for Computational Linguistics.

Longxiang Shen, Bowei Zou, Yu Hong, Guodong Zhou, Qiaoming Zhu, and AiTi Aw. 2019. Negative focus detection via contextual attention mechanism. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2251–2261, Hong Kong, China. Association for Computational Linguistics.

György Szarvas, Veronika Vincze, Richárd Farkas, and János Csirik. 2008. The bioscope corpus: Annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '08, pages 38–45, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ann Taylor, Mitchell Marcus, and Beatrice Santorini. 2003. *The Penn Treebank: An Overview*, pages 5–22. Springer Netherlands, Dordrecht.

Veronika Vincze, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9:S9 – S9.

Xin Wang, Yuanchao Liu, Chengjie Sun, Baoxun Wang, and Xiaolong Wang. 2015. Predicting polarities of tweets by composing word embeddings with long short-term memory. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1343–1353, Beijing, China. Association for Computational Linguistics.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68, Uppsala, Sweden. University of Antwerp.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.

Bowei Zou, Guodong Zhou, and Qiaoming Zhu. 2014. Negation focus identification with contextual discourse information. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 522–530, Baltimore, Maryland. Association for Computational Linguistics.

Bowei Zou, Guodong Zhou, and Qiaoming Zhu. 2015. Unsupervised negation focus identification with word-topic graph model. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1632–1636, Lisbon, Portugal. Association for Computational Linguistics.

A Appendix

In this section, we provide additional details on the neural models discussed in this paper.

A.1 Details on the Neural Architecture

The neural model shown in Figure 1 is our full model. It consists of a base network and additional components indicated with dotted shapes. The additional components incorporate information about the scope of the negation and context (previous and next sentence). In this section, we provide additional information about the input representation and the additional components.

Input Representation. As discussed in Section 4, we map each word token to its 1,024-dimensional pre-trained ELMo embedding (Peters et al., 2018). We do not update the ELMo embeddings during the training of the network.

Our baseline model leverages two additional embeddings for encoding positional information of the negated verb as well as the semantic role labels of the input tokens. We extract semantic roles from the training, development and test sets in the the the CoNLL-2005 Shared Task (Carreras and Màrquez, 2005). The embeddings indicating the negated verb and semantic role labels are trained from scratch along with all the other weights in the full network.

We employ an additional embedding to incorporate scope information into the network (Section 4). Like the two additional embeddings described above, the embeddings to indicate scope information are trained from scratch.

Figure 2 shows the construction of the input representation. The input sentence is "The carrier has not yet turned a profit." E_{token} (top) denotes the 1,024-dimensional ELMo embeddings of *token*. The other embedding vectors shown in Figure 2 are to indicate the position of the negated verb, semantic roles and the scope of the negation. We have two tags to indicate the negated verb ("Y" when the token is the negated verb and "N" otherwise), two tags to indicate the scope ("I_S" when the token is inside the scope and "O_S" otherwise), and 15 tags to indicate semantic roles (one per role label). All the embedding weights for each token are concatenated before feeding them into the first BiLSTM layer. The final input dimension per token is 1,474:1,024 from the word token embedding, 50 from the negated verb embedding, 200 from the semantic role embedding, and another 200 from the scope embedding.

Note that in the sample sentence shown in Figure 2, all tokens are inside the scope of the negation except the negation cue (negation cues are annotated as outside of the scope in CD-SCO (Morante and Daelemans, 2012)). The scope of a negation, however, can span over all the tokens or a small part of a sentence, or even be discontinuous (Morante and Daelemans, 2012). In the example sentence below, for example, the scope of the negation only spans over the last clause:

> Mr./O_S Paul/O_S says/O_S he/O_S had/O_S not/O_S one/O_S but/O_S four/O_S advisers/O_S and/O_S that/O_S he/I_S never/O_S bid/I_S impulsively/I_S ./O_S

BiLSTM-Attention Network for Context. To capture contextual information, we add two attention-based recurrent networks, one for the previous sentence and another one for the next sentence. These additional components are shown in
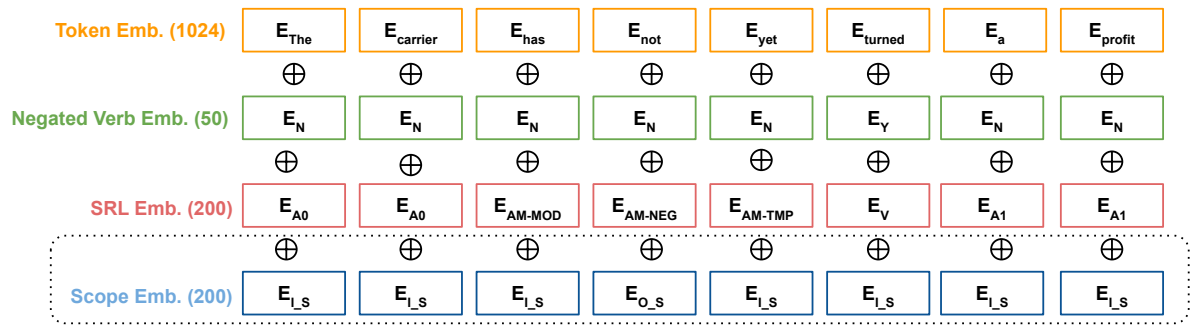
Token Emb. (1024)	E_{The}	$E_{carrier}$	E_{has}	E_{not}	E_{yet}	E_{turned}	E_a	E_{profit}
	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Negated Verb Emb. (50)	E_N	E_N	E_N	E_N	E_N	E_Y	E_N	E_N
	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
SRL Emb. (200)	E_{A0}	E_{A0}	E_{AM-MOD}	E_{AM-NEG}	E_{AM-TMP}	E_V	E_{A1}	E_{A1}
	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Scope Emb. (200)	E_{I_S}	E_{I_S}	E_{I_S}	E_{O_S}	E_{I_S}	E_{I_S}	E_{I_S}	E_{I_S}

Figure 2: Input representation of our neural model. The symbol ⊕ denotes concatenation, not addition.

the left and right dotted rectangles in Figure 1. The previous and next sentences in the example shown in Figure 1 are "StatesWest operates four twin-engine turboprop aircraft, connecting 10 cities in California, Arizona and Nevada" and "The former president of FirstSouth F.A., a defunct Arkansas thrift, pleaded guilty to conspiring to inflate the institution's earnings by concealing worthless loan guarantees" respectively.

Like in the baseline model, we map each word of the adjacent sentences to its 1,024-dimensional ELMo embedding vector before feeding them into the recurrent network. Each network component consists of a 2-layer Bidirectional LSTM with 50 hidden units. A dropout rate of 30% is applied to the recurrent layers. We add an attention layer on top of the final BiLSTM layer. More specifically, we adopt the word-attention technique proposed by Yang et al. (2016). The attention weights from both networks are concatenated with the final hidden representation of the base 3-layer BiLSTM network (Figure 1). Subsequently, the additional network components are trained with the original BiLSTM network.